Supplementary Materials

Efficient prediction of potential energy surface and physical properties with Kolmogorov-Arnold Networks

Rui Wang[1,2], Hongyu Yu[1,2], Yang Zhong[1,2], Hongjun Xiang[1,2,*]

[1]Key Laboratory of Computational Physical Sciences (Ministry of Education), Institute of Computational Physical Sciences, State Key Laboratory of Surface Physics, and Department of Physics, Fudan University Shanghai, Shanghai 200433, China.
[2]Shanghai Qi Zhi Institute, Shanghai 200030, China.

[*]**Correspondence to:** Prof. Hongjun Xiang, Key Laboratory of Computational Physical Sciences (Ministry of Education), Institute of Computational Physical Sciences, State Key Laboratory of Surface Physics, and Department of Physics, Fudan University, 220 Handan Rd., Shanghai 200433, China. E-mail: hxiang@fudan.edu.cn

DETAILS FOR THE NETWORKS

Details for the allegro model

For the Allegro model[1] using MLPs, the two-body latent MLP consists of two layers of dimensions[16,32], using SiLU nonlinearities. The latent MLP consists of one layer of dimension 32, also using a SiLU nonlinearity. The output block MLP has one layer of dimension 32 and no nonlinearity.

For the Allegro model using B-spline basis[2], the two-body latent part consists of two layers of KANs of dimensions[5,10], using SiLU nonlinearities. The latent "MLP" part consists of one layer of KAN of dimension 10, also using a SiLU nonlinearity. The output block has one hidden layer of KAN of dimension 10 and SiLU nonlinearity. For all the KAN networks, the spline grid size is set to 5 and the spline order is set to 3.

For the Allegro model using Gaussian and Fourier basis[3], the two-body latent part consists of two layers of KANs of dimensions[5,10], using SiLU nonlinearities. The

latent “MLP” part consists of one layer of KAN of dimension 10, also using a SiLU nonlinearity. The output block has one hidden layer of KAN of dimension 10 and SiLU nonlinearity. For all the KAN networks, the number of grids is set to 8.

When replacing MLPs from different parts of the Allegro model with KANs, the part using MLPs has the same parameters as the Allegro model using MLPs, and the parts using KANs has the same parameters as the Allegro model using KANs.

The Allegro model using MLPs has 48,840, while the Allegro model using KANs with Gaussian basis functions contains 55,822 trainable weights. This increase is due to the more complex structure of KANs, which enhances the model's flexibility and accuracy but at the cost of longer training times. When KANs with Gaussian basis functions are used exclusively in the output block, the number of trainable weights is reduced to 53,849. This configuration leverages the expressive power of KANs while limiting parameter growth in the two-body latent embedding and latent MLP part, striking a balance between computational efficiency and predictive accuracy. Consequently, the Allegro model using KANs in the output block achieves both higher accuracy and shorter training times compared to the models using only MLPs or KANs.

Details for the Neural Equivariant Interatomic potentials (NequIP) model

We only replaced MLPs form the “output_hidden_to_scalar” part in the NequIP model[4] by KANs. Both the MLP and the KAN in the “output_hidden_to_scalar” part is a linear layer. The parameters of the KAN networks are the same as with the Allegro model.

Details for edge-based tensor prediction graph neural network (ETGNN) model

We only replaced the up projection part of the node output block and update block in the ETGNN model[5]. Both the MLP and the KAN in the “output_hidden_to_scalar” part is a linear layer. The parameters of the KAN networks are the same as with the Allegro model.

DETAILS FOR THE TRAINING DATASETS

Details for the Ag dataset

The Ag dataset[1] was generated from ab-initio molecular dynamics (AIMD) simulations of a bulk face-centered cubic structure containing a vacancy and comprising 71 atoms. The AIMD simulations were conducted at a temperature of 1111 K, corresponding to 90% of the melting point of Ag. Frames were extracted at intervals of at least 25 fs to minimize correlation within the trajectory, and each frame was recalculated using converged density functional theory (DFT) parameters. The simulations were carried out with the Vienna Ab-Initio Simulation Package (VASP)[6], employing the PBE exchange-correlation functional[7] and a cutoff energy of 520 eV, with Gamma-point k-point sampling. The dataset consists of 1,000 unique structures, of which 950 were used for training and 50 were used for validation.

Details for the HfO2 dataset

The HfO_2 dataset[8] was created using density functional theory (DFT) calculations conducted with the VASP package[6]. Initial structures were generated by randomly perturbing the ground-state configurations of the $P2_1/c$, $Pbca$, $Pca2_1$, and $P4_2/nmc$ phases of HfO_2. These structures were then sampled through NPT simulations across a range of temperatures (100 to 3300 K) and pressures (-50 to 400 kBar). The projected augmented wave method[9,10] was utilized, along with the PBE exchange-correlation functional[7]. An energy cutoff of 600 eV was applied. From this dataset, we selected 10,000 structures, with 9,000 used for training and 1,000 for validation.

Details for the SiO_2 dataset

The SiO_2 dataset[5] consists of 3,992 randomly perturbed SiO_2 structures calculated using density functional perturbation theory (DFPT). The dataset was split into training, validation, and test sets in a 6:2:2 ratio. We calculated the Born effective charges using ETGNN with MLPs and KANs with Gaussian and B-spline basis functions.

Details for the LAMMPS simulation

We assessed the inference speeds and GPU memory usage of different models by running molecular dynamics simulations with the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS)[11]. These simulations utilized the Allegro pair style implemented in the Allegro interface[1]. The initial structure was obtained

from the Ag dataset[1]. Simulations were carried out under an NVT ensemble at 300 K, with a time step of 1 ps. For each model, we conducted 5,000 time steps to evaluate inference speed.

REFERENCES

1. Musaelian A, Batzner S, Johansson A, Sun L, Owen CJ, Kornbluth M, Kozinsky B. Learning local equivariant representations for large-scale atomistic dynamics. Nat Commun. 2023 Feb 3;14(1):579.
2. Blealtan. Blealtan/efficient-kan [Internet]. 2024 [cited 2024 Aug 15]. Available from: https://github.com/Blealtan/efficient-kan
3. Li Z. Kolmogorov-Arnold Networks are Radial Basis Function Networks [Internet]. arXiv; 2024 [cited 2024 Aug 11]. Available from: http://arxiv.org/abs/2405.06721
4. Batzner S, Musaelian A, Sun L, Geiger M, Mailoa JP, Kornbluth M, Molinari N, Smidt TE, Kozinsky B. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. Nat Commun. 2022 May 4;13(1):2453.
5. Zhong Y, Yu H, Gong X, Xiang H. A General Tensor Prediction Framework Based on Graph Neural Networks. J Phys Chem Lett. 2023 Jul 20;14(28):6339–48.
6. Kresse G, Hafner J. Ab initio molecular dynamics for liquid metals. Phys Rev B. 1993 Jan 1;47(1):558–61.
7. Perdew JP, Burke K, Ernzerhof M. Generalized Gradient Approximation Made Simple [Phys. Rev. Lett. 77, 3865 (1996)]. Phys Rev Lett. 1997 Feb 17;78(7):1396–1396.
8. Wu J. Deep learning of accurate force field of ferroelectric HfO_2. Phys Rev B. 2021;103(2).
9. Blöchl PE. Projector augmented-wave method. Phys Rev B. 1994 Dec 15;50(24):17953–79.
10. Kresse G, Joubert D. From ultrasoft pseudopotentials to the projector augmented-wave method. Phys Rev B. 1999 Jan 15;59(3):1758–75.
11. Thompson AP, Aktulga HM, Berger R, Bolintineanu DS, Brown WM, Crozier PS, in 't Veld PJ, Kohlmeyer A, Moore SG, Nguyen TD, Shan R, Stevens MJ, Tranchida J, Trott C, Plimpton SJ. LAMMPS - a flexible simulation tool for particle-based materials

modeling at the atomic, meso, and continuum scales. Comput Phys Commun. 2022 Feb 1;271:108171.